African Journal of Biological Sciences

PROBABILISTIC NEURAL NETWORK APPROACH FOR LUNG CANCER DETECTION

B. Usha Priya[a*] & Dr. V. Lokeswara Reddy[b]

1 Research Scholar, Department of CSE, JNTUA, Ananthapuramu, A.P India.
2 Professor, KSRM college of Engineering, CSE Department, A.P India.
ushapriya512@gmail.com , vlreddy74@gmail.com

Abstract: Lung cancer is among the most significant fatal cancers, and according to statistics, it is the second disease-causing high mortality rate among women/men. Early detection can lead to survival with the proper treatment process. The detection of Lung cancer manually consumes money and time and is inefficient in several cases. In this proposed system, Lung cancer is segmented and detected using an artificial neural network classification method in mammogram images. "Lung cancer GUI"- A MATLAB-based Graphical User Interface (GUI) is created in the proposed system. MIAS database mammogram images are uploaded to the "Lung cancer GUI," and the image is preprocessed using the adaptive median filtering method to remove noises. Then the images are segmented and extracted features using the Gaussian mixture model (GMM) with a gray level co-occurrence matrix. These extracted features are trained and classified utilizing the probabilistic neural network (PNN) classification approach. The proposed system identified and classified the images into benign, malignant, or normal types with 96.67% accuracy, 97.0% precision, 96.0% sensitivity, and 97.98% specificity.

Keywords: Lung cancer GUI, Gaussian mixture model (GMM), Mammography, Probabilistic Neural Network (PNN).

1. INTRODUCTION

The second leading cause of death in all is Lung cancer, amongst the most lethal forms of the illness. Women/men all over the globe are at danger of being detected with Lung cancer at some point in their life. Roughly 2.1 million women/men are diagnosed with Lung cancer each year, and it is predicted that 627,000 women/men deaths from the disease in 2018, accounting for around 15percent among all carcinoma fatalities amongst women/men worldwide. When compared to less developed places, the incidence of Lung cancer is greater in more advanced parts of the globe, and percentages are rising rapidly in every part of the globe [1]. By the year of 2030, it is estimated by the World Health Organization (WHO) that nearly 27 million novel Lung cancers cases will be diagnosed in women/men worldwide. Based on a statistic by Globocan 2018, there were 7593 female Lung cancer cases registered in Malaysia, which

accounting for 32.7% of all cancer cases registered and it ranks no.1 cancer case among Malaysian women/men in 2018[2].

There are many distinct forms of Lung cancer, and they may start in a variety of different places in the Lung, including the ducts, lobules, and even the tissues somewhere between. Lung cancer is often divided into two types: estrogen receptor positive and estrogen receptor negative.

- Non-invasive cancer: "In Situ" means "in situ," which implies that the abnormal cells stay in the specific location of the Lung (lobules or ducts) and do not migrate to adjacent tissue. Lung cancer: LCIS (Lobular Carcinoma in Situ) and DCIS (Ductal Carcinoma in Situ) are the two forms of carcinomas found in situ.

- Invasive cancer: Cancer that has spread throughout the body, The metastasis is not the same as the primary tumor. Metastasis is the term used to describe the process by which cancer cells break out from the main tumor and spread to other regions of the system via the blood or lymph vessels. Invasive cancer refers to cancer cells that have moved outside of the Lung's ducts or lobules and into the surrounding tissue. It is the most common kind of cancer in women/men. In certain cases, it has the potential to spread to other parts of the human body. This happens via blood arteries like lymphatic and veins vessels. Invasive Lobular Carcinoma (ILC) and Invasive Ductal Carcinoma (IDC) are the two kinds of invasive Lung cancer that may occur [3].

Most common ones are DCIS and IDC [4]. There are some less common ones like Medullary Carcinoma, Paget's Disease of the Nipple, Mucinous Carcinoma, Tubular carcinoma and Phyllodes tumor. Lung cancer may also be categorized utilizing genomic research, that will bring in a more exact categorization of Lung cancers depending on the proteins and genes that are expressed in the tumor cells.

Although no one understands what causes Lung cancer, there are a few risk factors that have been linked to an increased risk of Lung cancer in women/men. Some of these risk factors are uncontrollable, while others are connected to cancer-causing substances in the environment or personal decisions like pregnancy, smoking, or drinking. According to physicians, DNA mutations handed down through generations in a family cause 5 to 10% of Lung cancers. Lung cancer genes 1 and 2 are well-known Lung cancer genes that may be passed down through the generations and considerably enhance the chance of acquiring the disease [5]. Radiation exposure, starting periods at a younger age, starting menopause at an older age, postmenopausal hormone therapy, not having children or late pregnancies, oral contraceptive use, alcohol consumption, physical inactivity, obesity, and high-fat diets are all risk factors that can make women/men more likely to be diagnosed with Lung cancer. If Lung cancer is detected in its early stages, it is usually treatable and curable. If a tumor is found before it grows to a size of 10mm, the patient has an 85% probability of achieving full remission.

Early detection can lead to the survival with proper treatment process. Amongst the most common conventional methods utilized to identify and diagnosis of the Lung cancer at the earliest stage is mammography. Mammography, also known as mastography, is a diagnostic and screening procedure that involves the use of low-energy X-rays to examine the Lung [6]. Ionizing radiation will be used in mammograms to create images that will be evaluated by specialists for abnormal results. In this thesis, Mammography images are used which are obtained from open access database. The main databases which used are the "Mammographic Image Analysis Society" (MIAS) database and the "Digital Database for Screening Mammography" (DDSM). Figure 1(a) shows the normal Lung mammogram image which does not have any tumors and Figure 1(b) shows the abnormal Lung mammogram image which has tumors.
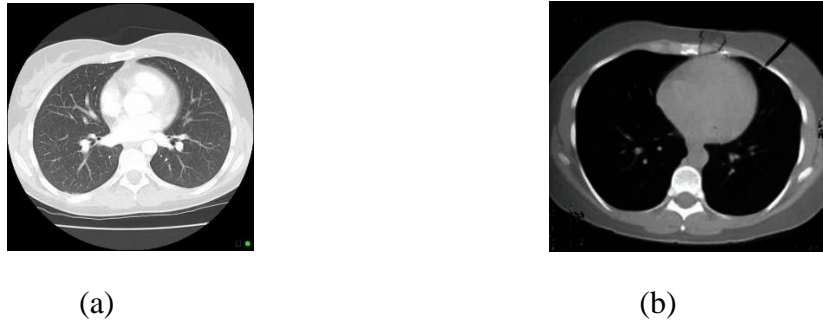
(a)   (b)

Figure 1: Lung Mammogram Images (a) Normal Lung (b) Cancerous Lung

The mammography process is uncomfortable for most women/men as some find it painful although this process takes only a few moments. One thing we have to know is that technologist cannot tell the patient the results of the mammogram. Only Lung specialist like surgical oncologist, Lung surgeon or radiologist can read the mammogram. He or she will look at the mammogram images for any early signs of Lung cancer. As this is not an automated or computer aided diagnosis, there is several disadvantages in the mammography method. Continuous efforts are in progress by researchers to create more effective computer aided ways for early and accurate diagnosis. Classical methods require Lung specialists to examine the Lung for detection and classification of various types of cancers are time and money consuming, sometimes become inefficient in several cases. And some of the other methods, like biopsy is a diagnosis procedure considered complementing the opinion of the pathologist. Thus, there is a need for more effective way to detect the Lung cancer without human involvement with high accuracy rate. In this thesis, a computer aided system is proposed which in the Lung cancer will be detected and segmented by a MATLAB based Graphical User Interface (GUI) which works based on artificial neural network (ANN) algorithm. This Computer Aided Diagnosis (CAD) system will be a second tool for Lung cancer specialists for discovering masses in the X-ray mammograms and distinguish the type of cancer developed in a patient.

1.1.Mammography

It is among the specialist health imaging methods that implements lower-energy X-rays (typically approximately 30 kVp) to analyze women/men's Lung tissue in order to recognize Lung cancer at initial stages. During a mammography exam, the inside anatomy of the Lung is shown, which is referred to as a mammogram. Radiography (often known as x-ray) is a non-invasive medical diagnostic that assists clinicians in the detection and treating variety of medical disorders. When imaging using x-rays, a tiny section of the body is exposed to a low dosage of ionizing radiation to have the pictures of the interior of the body, this is known as radiography [7]. The mammography technique has become the most commonly used method to detect Lung cancer. Today, specialized Lung X-ray equipment, which is used only for Lung X-rays, generates studies that are good in quality while using less radiation.

1.1.1. Procedure of Mammography

Mammography is a technique that may be performed as an outpatient. When the patient's Lung is placed in the mammography unit, a properly qualified radiologist/technologist will perform the procedure on the patient. Once the Lung has been put on a specially designed platform, it will be squeezed with a transparent plastic paddle. The technician will progressively pressure the patient's Lung over a period of time. Lung compression is necessary for a variety of reasons, including:

- Ensure that the Lung thickness is evened out so that all of the tissue can be seen.
- Spread out the tissue so that tiny irregularities are less likely to be covered up by covering Lung tissue after the procedure.
- As a result, a lower x-ray dosage may be used since a smaller area of Lung tissue is being scanned.
- Keep the Lung as steady as possible to reduce the amount of blurring produced by movement in the shot.
- Decrease x-ray scattering in order to improve image clarity.

During the time between images, the patient will be asked to shift positions. The two most prevalent sorts of views are those that look from the top to the bottom and those that provide an angled side perspective. The procedure will be performed on the Lung on the opposite side to ensure a perfect fit. During the x-ray technique, the patient must remain perfectly immobile and may be asked to hold their breath for a few seconds at a time in order to reduce the probability of a hazy image being taken. Depending on the scenario, the technician will walk behind a wall or into the next room in order to operate the x-ray equipment. In certain cases, once the examination is done, the patient may be asked to wait in his or her examination room till the radiologist determines that all relevant images have been taken. No more than thirty minutes should be required for the whole examination process.

This project is mainly focusing on develop a computer aided system to analyze the mammogram images in an automated manner with high accuracy rate. Currently, the mammogram images will be read by the Lung specialists to detect the abnormalities in the Lung which not an effective method as there are possibilities for error to occur. Hence, there is a need for develop like this CAD system for detection and diagnosis of cancers more accurately at early stage [8]. The scope of this project entails dividing the project into three stages.

The first stage is about to understand the concept of medical image processing. It is the technique and process used to get a picture of the inside structures of a person's body, which may be used for clinical diagnosis and medical intervention in order to learn more about certain disorders or the operation of certain organs or tissues. The interior parts of the body that are buried beneath the skin and bones will be exposed via medical imaging, which will aid in the identification and treatment of any disorders that may exist. Medical imaging also creates a database of regular anatomy and physiology of human body which will be easy to identify the abnormalities. The medical imaging procedures will vary according to the parts of the human body and also depend on the need of a doctor in detection of a health issue. For example, Magnetic Resonance Imaging (MRI) is best method for a physician if want study about organs with soft tissues like brain, spinal cord and heart. X-rays is the most suitable medical imaging technique if want identify and study about hard tissues in human body like bone, tooth enamel, dentin, cementum and cartilage [9]. The difference in medical imaging not only due to the difference part of the body, but also involve the different method of processing that image. Medical imaging started to develop quickly now because of the rapid development of image processing techniques which increase the percentage accuracy in detecting a disease. This project mainly focuses on mammography which is specialized medical imaging procedure in early detection and diagnosis of Lung cancer in women/men. Mammography uses a low-dose x-ray system to create visual of interior parts of the Lungs. The obtain X-ray image of the Lung is called mammogram.

The CAD system is a method based on complex computer algorithm which can be created specifically to analyze some complicated structures in medical images and develop a decision-making system for those images based on the structures, thus can identify the disease. Within structure of this project, the CAD procedures followed in tumor detection in the mammogram

images are the pre-processing, usage of adaptive median filter, image segmentation, features extraction and classification the type of cancer using artificial neural network algorithm.

The second stage of this project is about the development and implementation of a graphic user interface (GUI) based on MATLAB platform. The codes for the GUI will be written and developed in the MATLAB GUI environment. For the digital image processing technique in analyzing the mammogram image, the MATLAB image processing tool is used to write the codes together with the GUI codes with the help of existing algorithms from some online open source. Artificial neural network algorithm is used to classify the tumor. The codes must be able to segment the tumor and classify the mammogram image as benign (non-cancerous) or malignant (cancerous) tumor.

2.   LITERATURE REVIEW

According to study conducted by Scott Mayer Mckinney and colleagues [10], it has been shown that an artificial intelligence (AI) system can outperform human cancer experts in the detection of Lung cancer. As part of this project, they assembled a large representative dataset from the United Kingdom and a large enriched dataset from the United States, with the goal of testing the model's performance in the clinical setting. Results show that absolute false positives have decreased by 5.7 percent and 1.2 percent in the US and the UK, respectively. Absolute false negatives have decreased by 9.4 percent and 2.7 percent, respectively, in both countries. When tested against six radiologists in an independent study, the artificial intelligence system outperformed every single human reader. After performing a thorough evaluation of the artificial intelligence system, the researchers will proceed with clinical investigations aimed at improving the accuracy and efficiency of Lung cancer screening in the future.

When it comes to identifying Lung cancer at an early stage, the mammography screening procedure is the most widely utilized tool currently available. However, this procedure is far from ideal, and the precision of the data is compromised as a consequence. As a result of the existence of unknown noise, low picture contrast, lack of homogeneity, weak borders, and unconnected portions, the content of the mammography images is impacted negatively. It is possible to resolve this problem by using pre-processing procedures. In their publication, Ramani et al. [11] discussed the need of pre-processing methods in order to fix and adapt the bad acquired mammogram picture before further research and processing can be performed. Pre-processing may be accomplished by the use of a variety of filtering methods. It is used to improve picture quality, reduce noise, retain image edges, enhance and smoothen the image by using the radial gradient filter. It was discovered that four forms of filtering were utilized in this research, with the means square error (MSE), peak signal-to-noise ratio (PSNR), structural content (SC) and normalized absolute error being the most important metrics to be determined (AE). They conducted their investigation using mammography pictures from the MIAS database. Their findings indicate that, when compared to other approaches, the adaptive median filter is the most appropriate way to utilize in pre-processing since the picture quality of the adaptive median filter is superior to that of the others.

It was stated in a study written by Anuj Kumar Singh and Bhupendra Gupta [12] that they established a quick and straightforward approach for detecting cancers in mammograms. Following the detection step, the tumor area in a mammography picture was segmented using image segmentation software. To find anomalies in images, they use straightforward image processing methods such as averaging and thresholding. The detection and segmentation procedures in their approach are the two most important elements in the process. During the detection phase, the original input picture is subjected to an averaging filter and thresholding operation, which are both performed simultaneously. They employed a rectangle window around the outputted area to locate malignant tissues, which they described in their work as the

Max-Mean and Least-Variance techniques, which they used to find malignant tissues. To segment a tumor patch during the segmentation stage, the patch is closed utilizing the morphological closure operation and the image gradient approach to determine the area border. Currently, the majority of their study is focused only on identifying the malignant tumor in the mammography picture.

Poonam Kathale and Snehal Thorat [13] had written a study in which they described a strategy for detecting cancer regions and categorizing patients as normal or malignant based on their findings. They employed mammography images taken from two different perspectives: Cranio-caudal (CC) and Mediolateral Oblique (MO) (MLO). The median filtering technique was used to do pre-processing on the input mammography picture in their suggested solution. The characteristics are extracted with the help of a Gray Level Co-occurrence Matrix. The Random Forest (RF) classifier was used to differentiate between cancer patients and healthy patients. This picture shows a tumor area split from a mammography image using a morphological technique, with the part highlighted in the original mammogram image. The accuracy of the RF classifier was 95.3 percent, the specificity was 83.33 percent, the sensitivity was 96.55 percent, and the precision was 98.25 percent for the experiment.

On the basis of an existing approach, Navdeep Kaur and Ajay [14] conducted research on the classification of normal and malignant cells using a weight-based algorithm. They then developed a novel methodology to overcome the limitations of the previous method. As a result of the usage of a Hidden Markov Model (HMM) classifier during the classification process, they discovered that the weight-based technique took a long time to categorize the data and that its accuracy was low. They recommended the use of a Bayesian classifier in the feature selection phase for three crucial variables: mass, density, and margin, in order to improve the accuracy of the classification. Increased identification rates may be achieved by expanding the classifier's capabilities to incorporate new characteristics such as tissue color. MATLAB is used to carry out the simulation, and it has been determined that the suggested approach outperforms the competition in terms of fault detection rate (accuracy), MSE, and PSNR.

In their paper, Jose Manuel Ortiz-Rodriguez et al. [15] stated that there is a need for more efficient methods to diagnose cancerous cells without the involvement of humans in order to obtain high accuracy results and overcome the limitations that oncologists face when using conventional methods in the detection of Lung cancer. They placed an emphasis on the usage of image processing methods and had employed them to generate imaging biomarkers via mammography analysis. They also relied on artificial intelligence technology in the hopes of detecting Lung cancer in its early stages. They employed a generalized regression artificial neural network to automate the categorization process (GRANN). The neural network was trained and tested to discriminate between malignant and benign tumors, and it was found to be 95.83 percent accurate in its classification. A CAD system based on the biomarker and a trained neural network is currently being developed.

A new method for automatic preprocessing image has been proposed by Abdelali Elmoufidi [16], who published an article in which he proposes a new method for preparing mammogram images before using them for further processing. It is mandatory to have a pre-processing method in preparing mammogram images before using them for further processing. Typically, mammography pictures include noise as well as other unrelated elements such as artifacts, labels, and so on. The suggested approach is divided into five parts, which are as follows: suppression of other irrelevant items, elimination of additional black background, determination of the mammography orientation, segmentation of the pectoral muscle, and, finally, augmentation of contrast. In order to validate this technique, it was evaluated on two databases: the DDSM database and the mini-MIAS database, which each include 10480 and

322 mammograms, respectively. Overall, the approach had reached an accuracy of around 99.5 percent.

Earlier this year, Than Than Htay and Su Su Maung [17] released research on the development of a system for identifying anomalies in mammograms that may be used to diagnose Lung cancer in its early stages. The median filter in the pre-processing step of their suggested technique was utilized to eliminate noise and undesired items from the mammography picture, according to the researchers. This is followed by Otsu's thresholding approach for Lung area segmentation on a non-uniform backdrop, which is described below. Following the thresholding of the pictures and the use of GLCM based texture analysis and feature extraction methods, the emphasis of feature extraction is on first order statistical analysis. The k-Nearest Neighbor (k-NN) classifier is used to determine if all of the collected characteristics are benign or malignant at the end of the process. The mini-MIAS database is used to assess the suggested method by the authors.

Using the Gaussian Mixture Model and Convolutional Neural Networks, Shamy and Dheeba [18] did research on the categorization of Lung cancer using these two methods. There are three steps to the technique that they have presented. The first step is to identify a ROI on the output image of the pre-processed picture. The ROI was discovered by the use of a K-means based GMM segmentation algorithm. After that, the ROI texture features are extracted and optimized using an optimal feature selection method, and the process is repeated until the desired result is achieved. Texture analysis based on GLCM is used to extract and compute the features. The selection is carried out using the Wrapper method/Hill Climbing algorithm, which is based on the Genetic Algorithm. The last step involves the categorization of tumors as benign or malignant, which is accomplished via the use of Convolutional Neural Network (CNN). The MIAS database was utilized to analyze this suggested approach, and it was found to be 95.8 percent accurate in its results.

Anwar Yahya [19] presented a novel paradigm for Lung cancer screening utilizing mammography pictures, which was accepted by the American Cancer Society. According to this research, an automated classification method for classifying Lung cancer into three types: normal, benign, and malignant; image enhancement; pectoral muscle separation; feature selection; and a classification procedure are all discussed in detail. This framework makes use of a variety of strategies to do all of this. To identify the particular characteristics in the mammography pictures, a novel approach called as Weighted Sparse Principal Component Analysis (WSPCA) is implemented. In this study, the performance of the Decision Tree employing SVM and Bayesian classification is tested utilizing the MIAS database, and the studied mammograms are categorized as benign or malignant based on the comparison. The suggested approach makes use of a Decision Tree classifier to determine classification. The validation was completed entirely with the help of the MIAS database. Utilizing the Decision Tree classifier, this framework obtained an overall accuracy of 90 percent, while 97.8 percent performed accuracy utilizing WSPCA features with the Decision Tree classifier for sequential selection of benign vs malignant mammograms was reached using the Decision Tree classifier.

Many different kinds of automated Lung cancer screening systems using mammography pictures have been investigated and evaluated over a long period of time, as seen in this review of the literature. This research has proved beneficial for future improvements in medical image processing with a high accuracy rate, which will be of great use. Some of the ways include the use of artificial intelligence in this industry in order to construct completely automated systems that do not need the participation of humans. Due to the difficulty in obtaining real patient data from healthcare facilities and the need for confidentiality, the majority of the studies relied on open-source databases such as the DDSM and MIAS databases for their study. For example, some researchers assessed current methodologies and conducted fresh research in order to get

a high rate of accuracy. As time passes, the intricacy of studies becomes more distinctive, and researchers offer new approaches that are becoming more difficult. Several key image processing approaches in Lung tumor identification were discussed in this literature review, and a few kinds of classifiers were employed to discriminate between benign and malignant tumors in mammography pictures were described in this literature review. This study provided additional in-depth information on image processing methods like pre-processing, segmentation, and features' selection that are used in the analysis of early Lung cancer.

Ibrahima sory keita and colleagues conducted a study in which they focused on the categorization of the brain tumor with the help of existing illness images and implementing planned mixture bright technique. Results indicated that the method used had an accuracy of 99% with a sensitivity and specificity of 99.14% and 98.79% respectively. These results indicated the efficiency and effectiveness of the selected method for identifying normal and non-standard tissues from Intelligent MRI [20].

S. Deivasigamani et al. conducted a study in which they detected the Glioma Brain Tumor with the help of Neural networks. Authors used the open access databases to gain the brain MRI images. They used the Gabor transform to extract the fetures of the image. Authors found that the sensitivity was 97.62%, specificity was 99.98%, positive predictive value was 99.06%, negative predictive value was 99.41% and accuracy was 99.40%. Authors concluded that proposed method is efficient and effective [21].

3. METHODOLOGY

3.1.Design:

The work divisions of Phase 1 and Phase 2 planned and scheduled are as shown in Figure 2. During the fourteen-week period, the workload is equally dispersed. The Gantt chart has been a very useful tool in ensuring that all of the activities are completed on schedule. Gantt Chart was used to construct the flowchart given below.
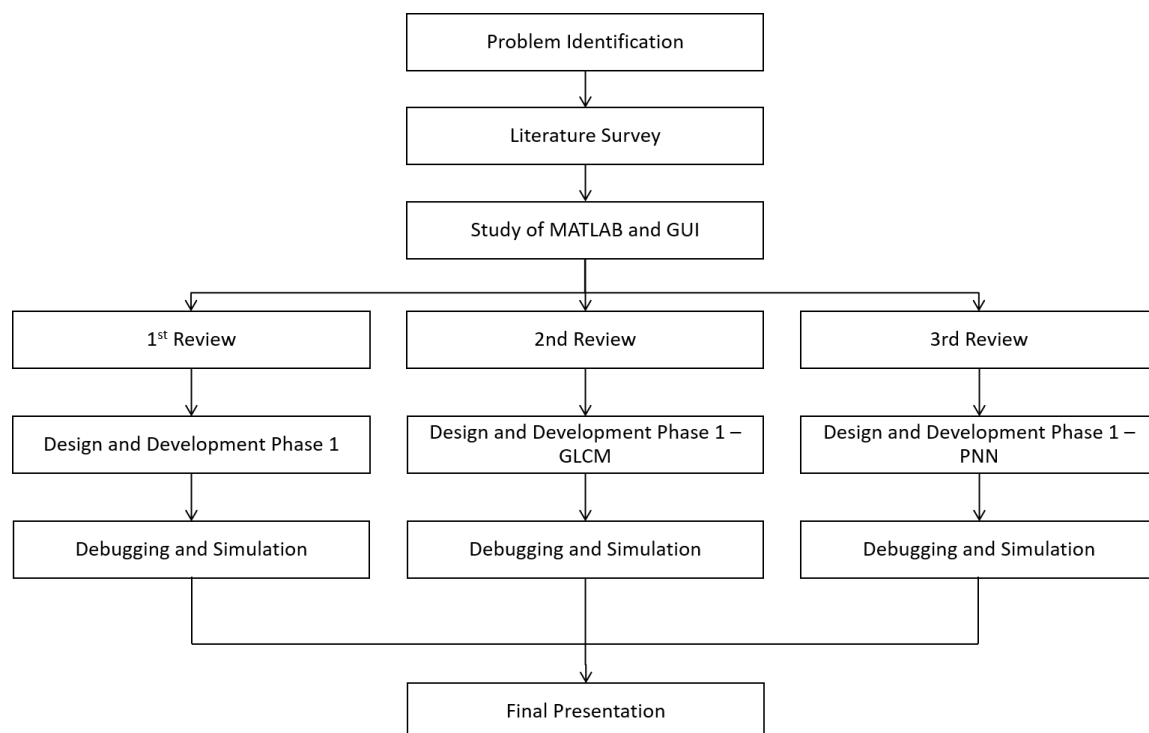
Figure 2: Representation of the flow of project work leading to the final representation.

In this project, a graphic user interface (GUI) is been developed in MATLAB GUI environment. The Lung mammogram images are taken from open access database which is MIAS database. The GUI used digital image processing methods to identify as well as segment the tumors in Lung. For classification of tumors, neural network algorithm is used. It classifies the tumor either as malignant or benign. The proposed method has the following steps in detecting the tumor: pre- processing, feature extraction GMM segmentation, and neural network (NN) classification. The proposed computer aided detection of tumor in a Lung and analysis methodology is shown in Figure 3.
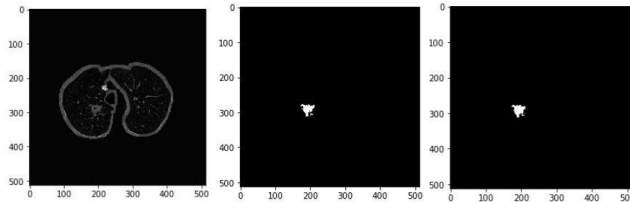
Figure 3. *U-Net image segmentation. Processed CT image (left), ground truth label (center), predicted label (right)*
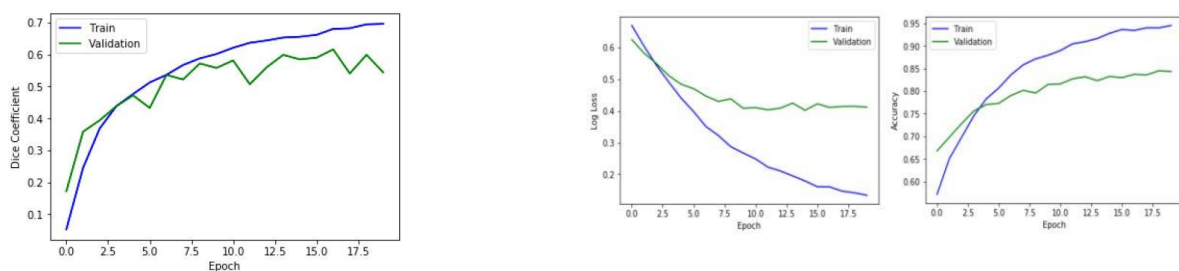
Figure 4: *CNN converges to a validation accuracy of 84.4% at classifying a detected nodule as TP or FP*

3.2. Instruments and Sample:

3.2.1. MIAS Database:

MIAS which is an UK research organization which interest in the studies of mammograms and had created a database of digital mammograms. It is most easily accessible and also most common databases to be used in Lung cancer research. It also known as mini-MIAS Mammographic database of mammograms.

3.2.2. Pre-processing: Adaptive Median Filter:

In image processing, the pre-processing step is the most important stage in the Lung mammography analysis because of the poor picture quality of the acquired mammogram in the first place. Pre-processing is important because it allows you to modify and alter the mammography picture before moving on to the next step. There are many different kinds of filtering methods that may be used for pre-processing. The adaptive median filter is one of the approaches used. This approach is used to increase picture quality, reduce noise, smoothen and enhance the image so that it may be utilized in a subsequent processing step.

In order to do this, pre-processing is performed in the orientation, augmentation, and segmentation of mammograms. The process might also include generating mask for pixels with the greatest intensity in order to lower resolution and partition the Lung. The process flow of

the suggested pre-processing approach for mammograms utilizing an adaptive median filter is shown in the following Figure 4, while Figure 5 show the sample outcome of the process flow using a mammogram image from mini-MIAS database.
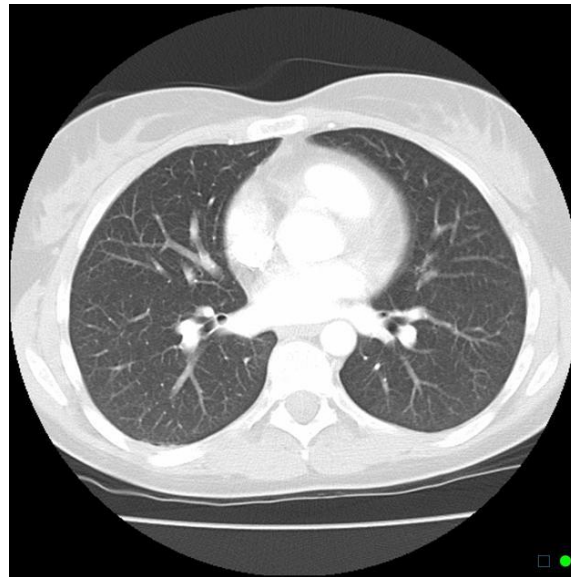

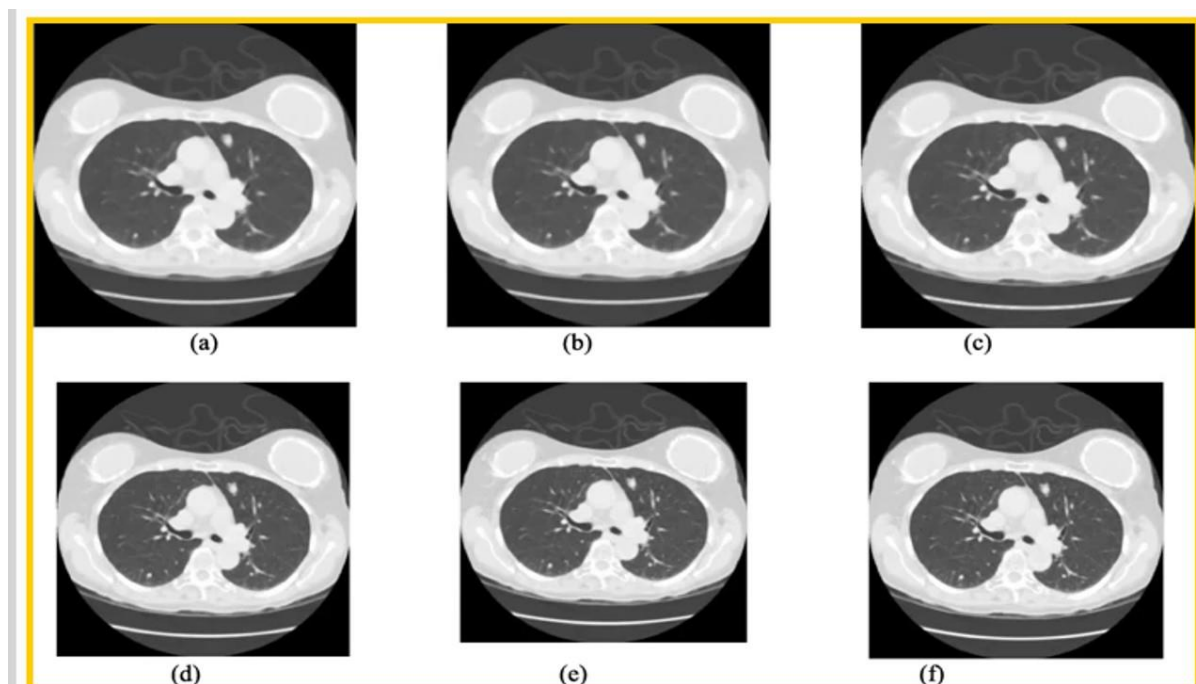**Figure 4: The process flow of proposed pre-processing method using adaptive median filter.**

Figure 5: The sample outcome of the proposed pre-processing algorithm.

3.2.3. *Gaussian Mixture Model (GMM) Segmentation*

The segmentation step is a critical and very significant procedure because all following steps such as features extraction and classification depend on its result. To extract the ROI from a mammogram image, the Gaussian Mixture Model (GMM) is suggested. GMM is a kind of clustering method that makes use of an iterative process called Expectation Maximization, which is very similar to k-means clustering. It belongs to the category of clustering algorithms.

3.2.4. Feature Extraction

The feature extraction concept is a component of the dimensionality reduction process, which involves dividing and condensing a large initial collection of data into smaller, more manageable groupings that can be identified from the input pattern. The presence of a huge set of variables in these big data sets is the most important aspect of these large data sets. The finest characteristics from huge datasets are extracted by a process of choosing and merging variables to form features, therefore significantly lowering the quantity of data. Following the description of the important properties of the picture into a feature space, the features extracted are presented to the classifier in order to display the features of the input patterns.

3.2.5. Neural Network Design

The classification stage plays a significant role in differentiating the normal mammogram from the abnormal mammogram. One of the well-known methods to classify the mammogram is using the Neural Networks (NN) based classifiers. An artificial neural network is a group of technologies that try to detect underlying links in a piece of data by mimicking the way the human brain functions. There are several types of neural networks which use different rules to learn and perform some process, but the pattern statistics from a collection of training datasets are analyzed in some manner by all of those neural networks, which then categorize new patterns based on the patterns that were previously classified.

3.3. Data Collection:

In this project, the mini – MIAS database is used for the segmentation and detection of tumor region in Lung images, then carry out the classification process. The modality of the Lung images in this database is mammogram. The database is based on mammograms which collected from the UK National Lung Screening Program. The mammography pictures from this freely accessible collection have been divided into two sets: a training set and a testing set, respectively. The training set has 20 normal pictures and 40 abnormal photos, whereas the testing set includes 50 normal pictures as well as 100 abnormal pictures, with the training set including 20 normal pictures and 40 abnormal pictures.

In this chapter, we present the created graphic user interface (GUI)- *'LungcancerGUI'*, and the step-by-step procedure to use it in the tumor detection process. A large number of simulations were carried out, and the results are shown in the following table. The characteristics taken from the preprocessed mammography picture using the GLCM approach were then evaluated with the PNN classification to identify the different kinds of tumors present. Finally, the detection of tumor by PNN classifier is shown in Figure 12. The *'LungcancerGUI'* had been tested with all types of tumors and shown in Figure 6 till Figure 14.

3.3.1. Graphic user interface (GUI)

- Defined as *'LungcancerGUI'* is designed and developed in the MATLAB GUIDE environment (Figure 6).
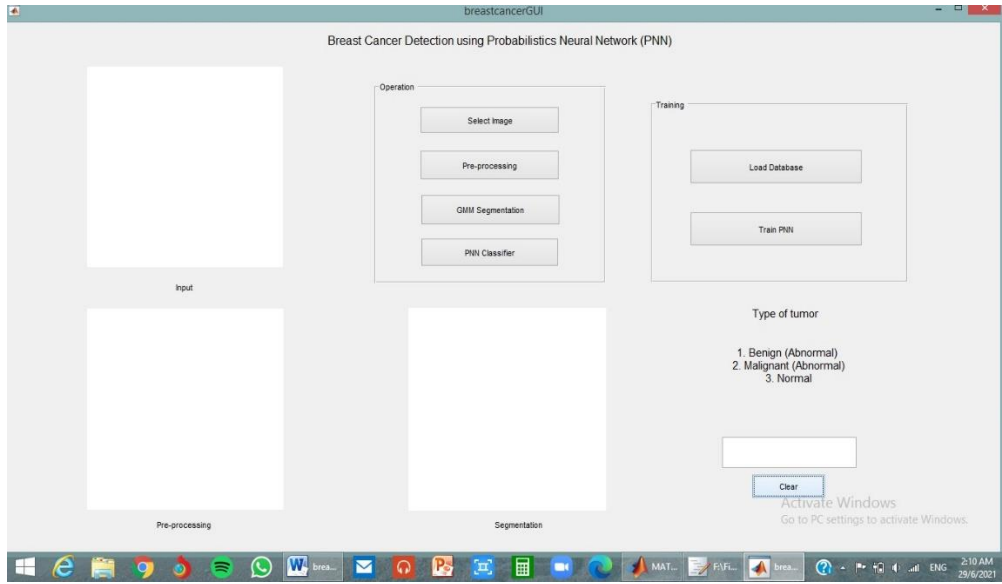
Figure 6: The Graphic User Interface – 'LungcancerGUI'

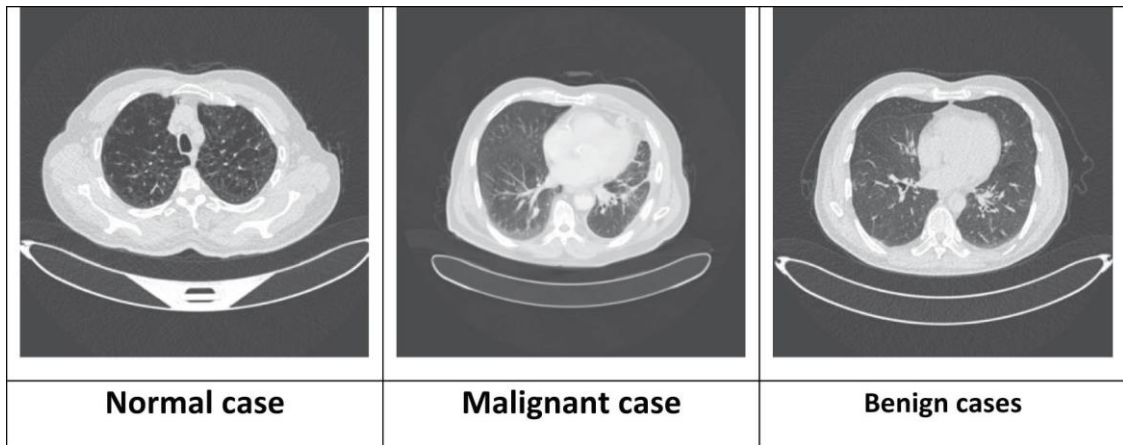- Browse and select mammogram image which extracted from mini-MIAS database (Figure 7).

Figure 7: Showing the 'Select image' window in the 'LungcancerGUI'.

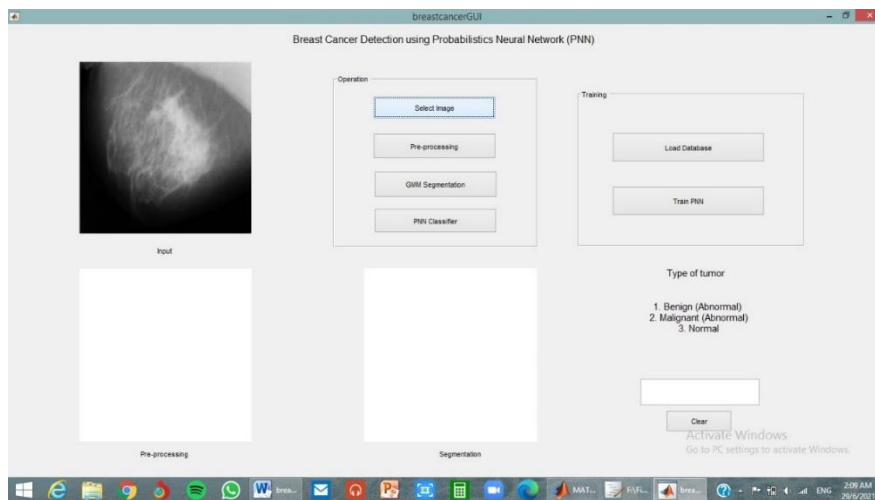- Upload the image to the GUI for further process (Figure 8).

Figure 8: Showing the Uploaded image to the GUI for further Process

3.3.2. PRE-processing stage –Adaptive median filter technique.

As stated before, the pre-processing is a significant stage in the Lung mammogram examination which is shown in Figure 9.
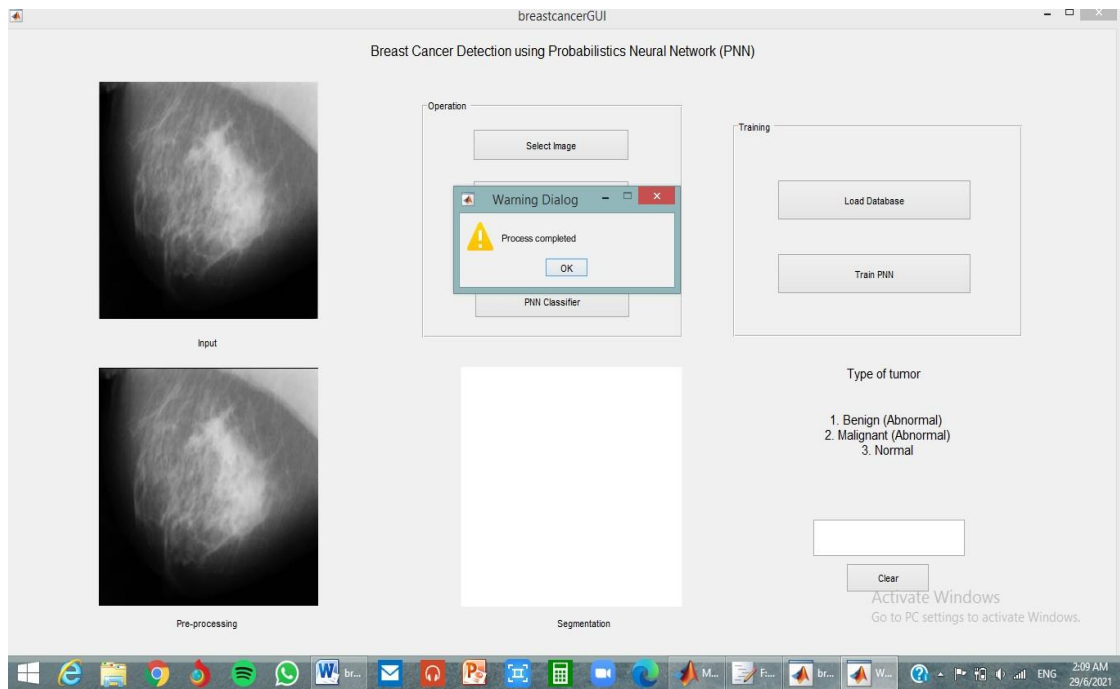
Figure 9: Input image and Output image after the pre-processing stage using adaptive median filter.

3.3.3. Segmentation stage – Gaussian Mixture Model (GMM) method

The segmentation step is a critical and very significant procedure because all following steps such as features extraction and classification depend on its result (Figure 10).
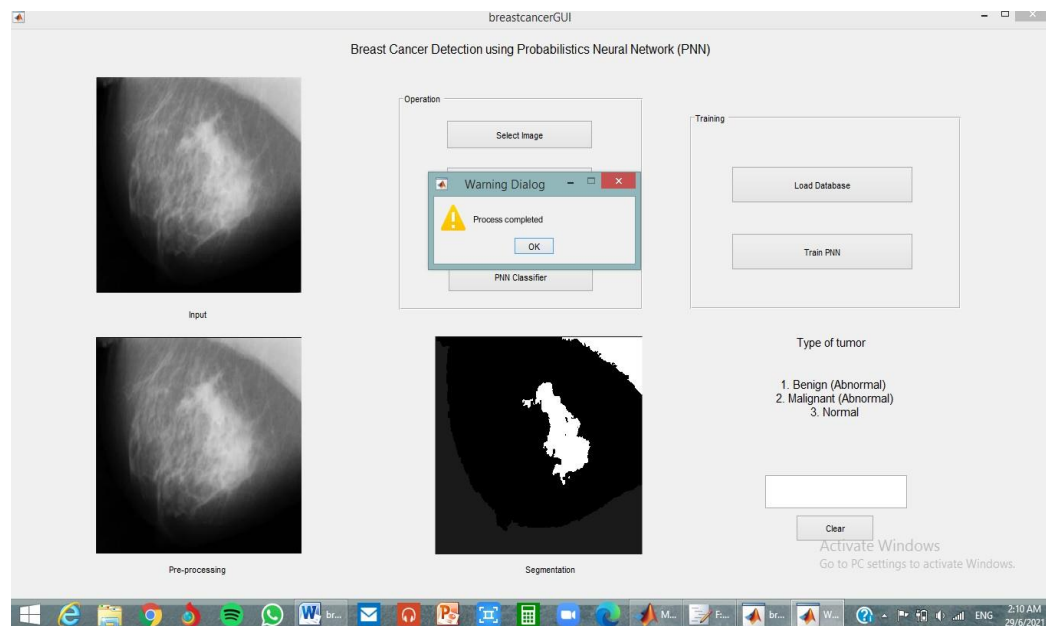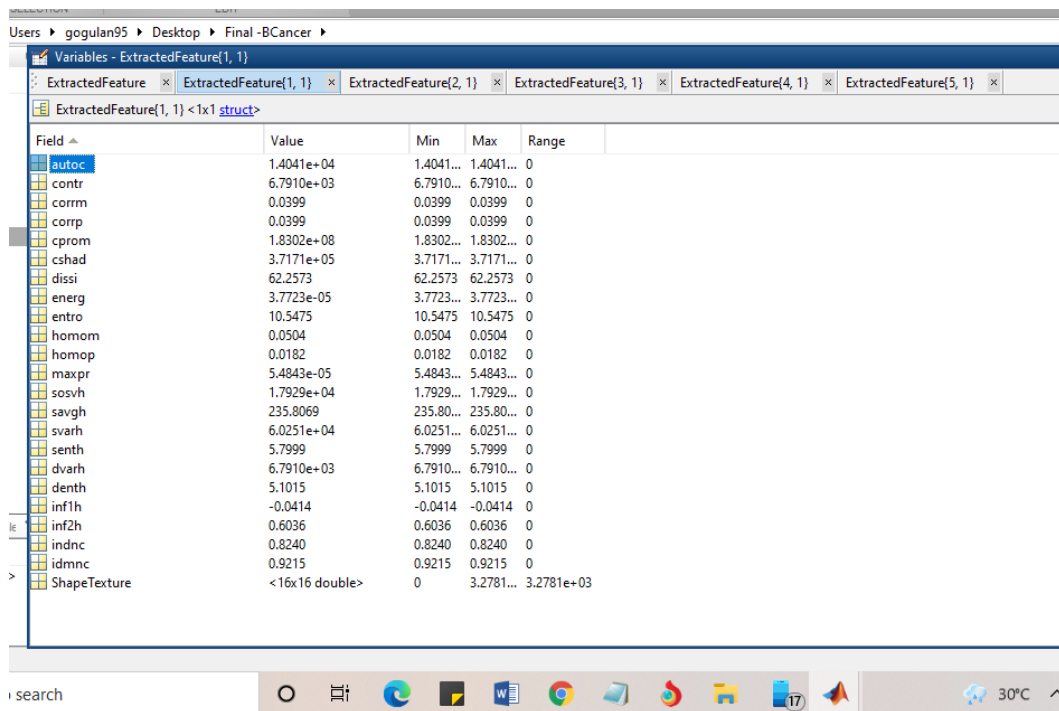
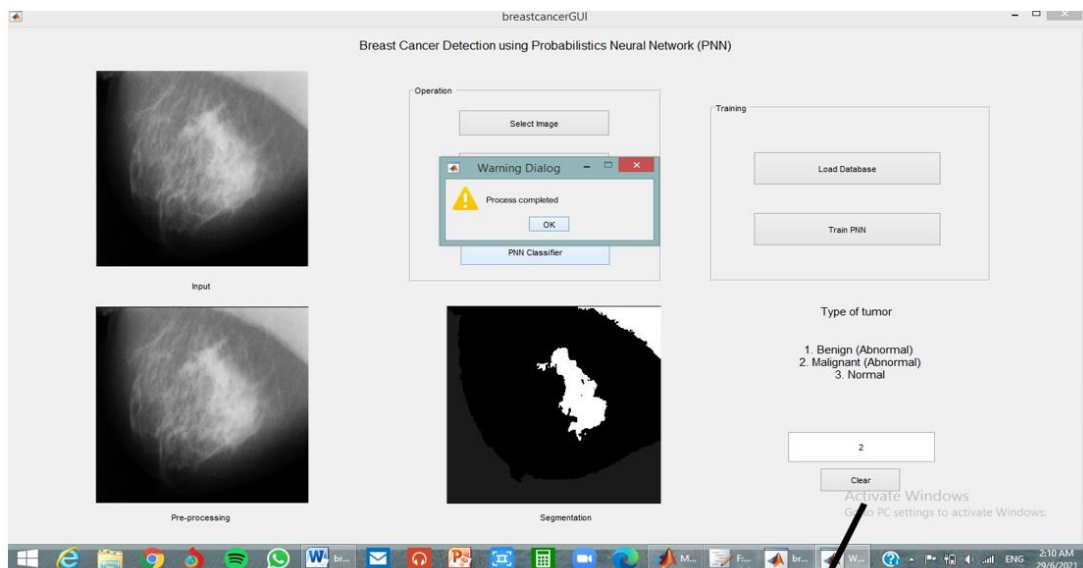**Figure 10: Output image after the segmentation stage by using Gaussian Mixture Model (GMM) method.**

B. Usha Priya/ Afr.J.Bio.Sc. 6(11)(2024).249-271

3.3.4. `Features Extraction stage – GLCM method`

Figure 11 is representing the Specific data of the selected image. Figure 12 is representing the final stage which was concluded with the help of PNN. Figure 13 and Figure 14 are representing the different types of tumors in the GUI and the process to clear all the input values.

Figure 11: Extracted features data of the specific mammogram image.

Final stage – Classification using Probabilistic Neural Network (PNN)

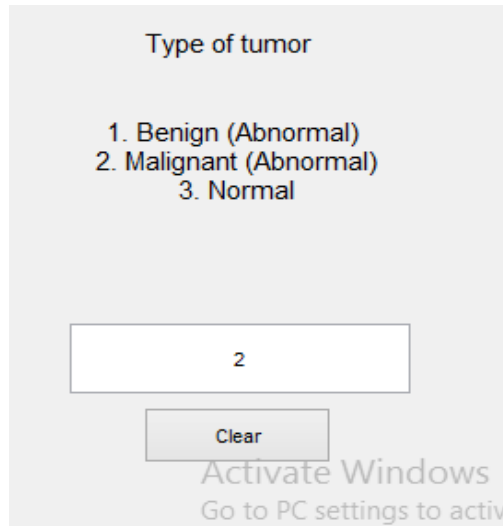**Figure 12: Classification of the Mammogram (Final Stage)**

Figure 13: Type of Tumor in the GUI which will be Shown as the Output.

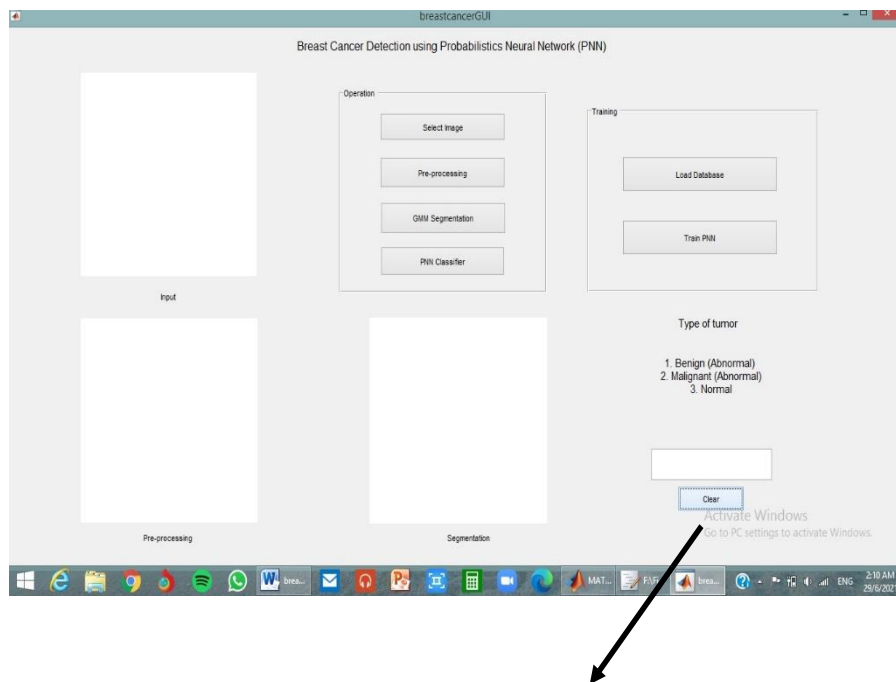**Figure 14: Press the 'Clear' button to reset the GUI for test another mammogram image.**

3.3.5. GUI test results using normal, malignant and benign type mammogram pictures.

GUI test results were gained for different tumor and normal Lung images with the help of normal, malignant and benign type mammogram pictures on the basis of comparison as shown below in Figure 15, Figure 16 and Figure 17.
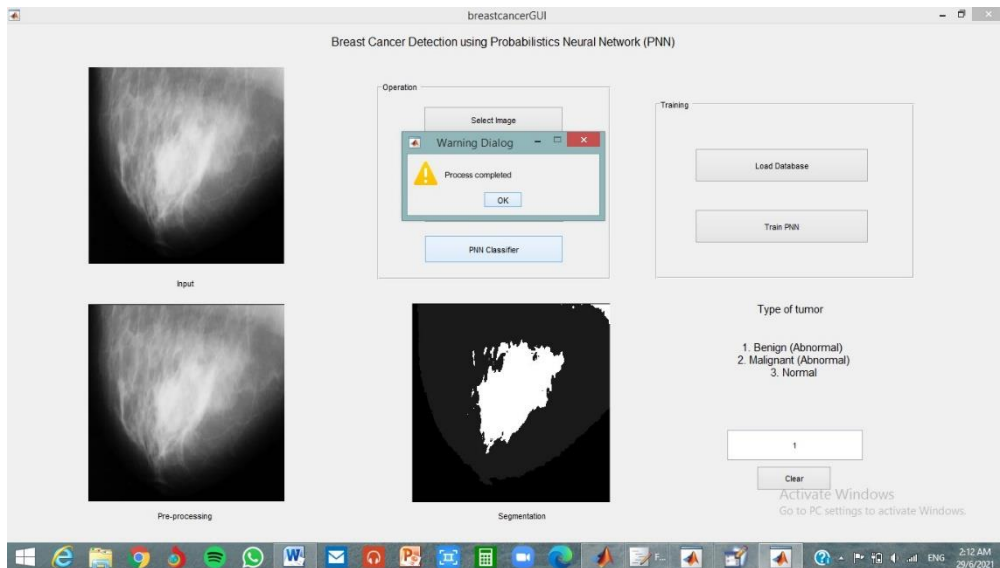
Figure 15: GUI test results of Benign tumor using normal, malignant and benign type mammogram pictures
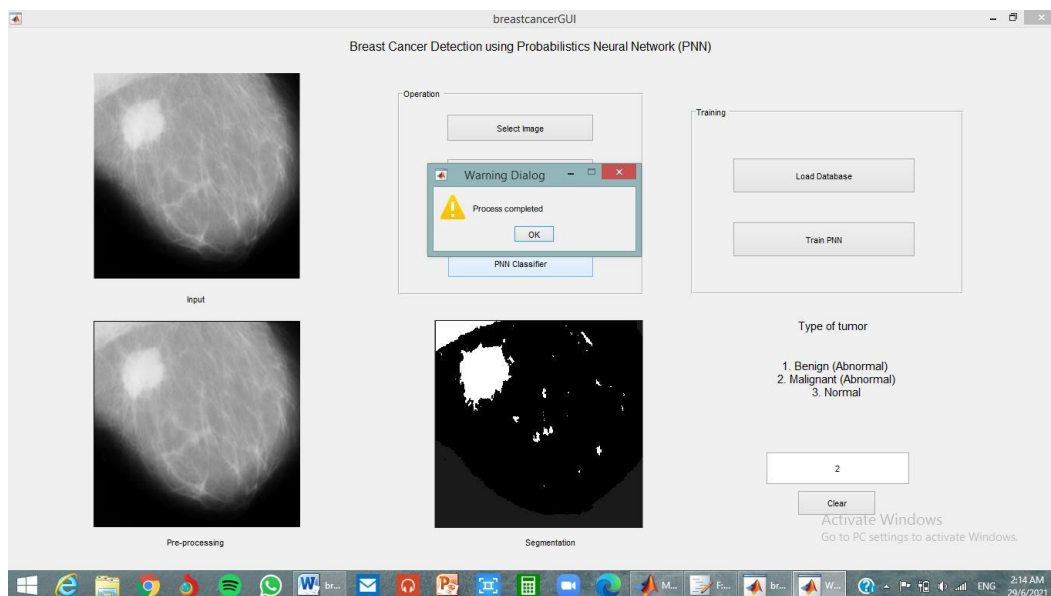

Figure 16: GUI test results of Malignant tumor using normal, malignant and benign type mammogram pictures
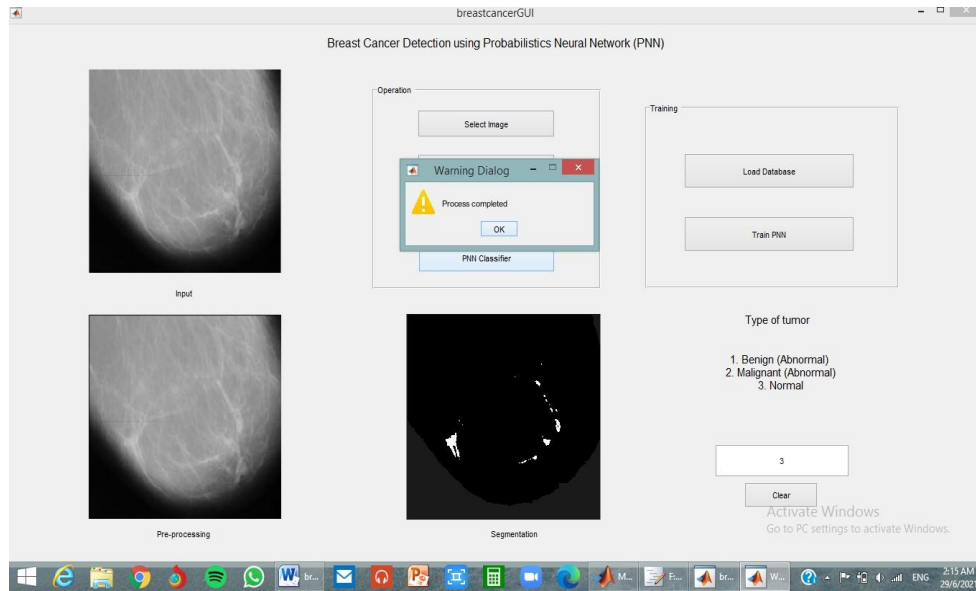
Figure 17: GUI test results of Normal / Healthy Lung using normal, malignant and benign type mammogram pictures

3.4.Data Analysis:

The performance of proposed method is analyzed with the following parameters:

- Specificity, Sp = T N / (T N + F P)

- Precision = T P / (T P + F P)

- Sensitivity, Se = T P / (T P + F N)

- Accuracy, Acc = (T P + T N) / (T P + F P + T N + F N)

These parameters are evaluated for a set image and listed in the Table 1 and Table 2 which "TP" stands for "True Positive"; "FP" stands for "False Positive"; "TN" stands for "True Negative" and "FN" stands for "False Negative". Accuracy is the ratio of total well- detected and classified the tumor, while precision is about the ratio of correctly predicted cases which actually turned out to be positive. Sensitivity, Se and specificity, Sp are about the ratio of well detected Lung cancer tumor and non-cancerous tumor. Table 1 and Table 2 evaluate the result of performance parameters for the proposed method by confusion matrix.

3.4.1.   Confusion matrix analysis for abnormal and normal tumor classification.

For this analysis, we categorized the tested 150 mammogram images results into abnormal and normal mammogram images.

Abnormal mammogram - Both benign tumor and malignant tumor categorized together in this analysis.

Normal mammogram – Healthy Lung images / Mammograms without tumor

Table 1: Confusion matrix for the abnormal and normal classification.

N=150	(Actual) Abnormal	(Actual) Normal	

(Predicted) Abnormal	TP = 97	FP = 2	**99**
(Predicted) Normal	FN = 3	TN = 48	**51**
	100	**50**	**150**

- TP: Abnormal Lung images correctly predicted as abnormal mammogram.
- TN: Normal Lung images correctly predicted as normal mammogram.
- FP: Normal Lung images incorrectly predicted as abnormal mammogram.
- FN: Abnormal Lung images incorrectly predicted as normal mammogram.

From this matrix, we can see that the PNN classifier made a total of 150 predictions. Out of 150 mammograms, the classifier predicted 97 mammograms as abnormal and 48 mammograms as normal. The actual data is 100 abnormal mammogram and 50 normal mammograms.

Table 2: Performance analysis for the abnormal and normal classification.

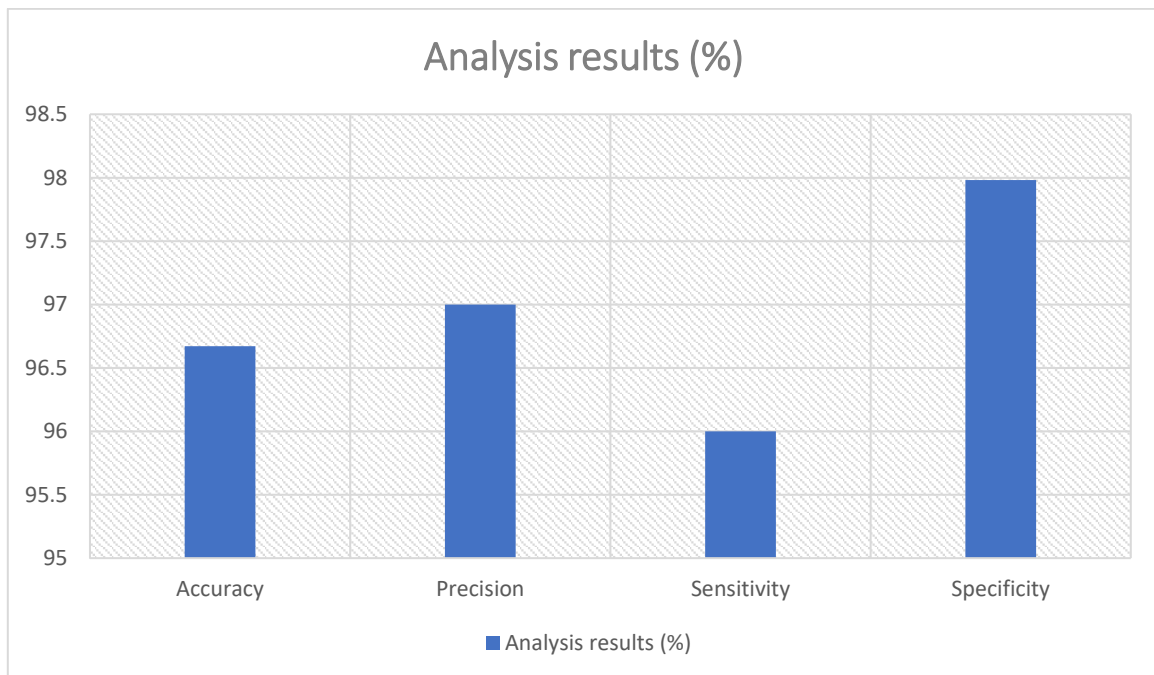| Performance Evaluation Parameters	Analysis results (%)
Accuracy	96.67
Precision	97.0
Sensitivity	96.0
Specificity	97.98

Figure 18: Graphical Analysis on Proposed Method on the Basis od Accuracy, Precision, Sensitivity and Specificity.

3.4.2. *Confusion matrix analysis for tumor types classification.*

For this analysis, we categorized the tested 150 mammogram images results as follows in Table 3:

Benign Tumor: Non-cancerous Malignant Tumor: Cancerous Normal: No tumors

Table 3: Classification of Tumor Types by Confusion matrix.

N=150	(Actual) Benign Tumor	(Actual) Malignant Tumor	(Actual) Normal	
(Predicted) Benign Tumor	TP = 49	1	0	**50**
(Predicted) Malignant Tumor	2	TP = 48	0	**50**
(Predicted) Normal	1	1	TP = 48	**50**
	52	**50**	**48**	**150**

Out of 150 mammogram, the classifier predicted 49 mammogram images as benign tumor, 48 as malignant tumor and another 48 mammogram as normal. The actual data is 50 mammogram images for each benign tumor, malignant tumor and normal.

Table 4: Performance Analysis for the Tumor Types Classification.

Performance Evaluation Parameters	Analysis results (%)
Accuracy	96.67
Precision **- Benign** **- Malignant** **- Normal**	 98.0 96.0 96.0
Sensitivity **- Benign** **- Malignant** **- Normal**	 94.2 97.96 100.0
Specificity **- Benign** **- Malignant** **- Normal**	 98.98 98.0 98.04

4. RESULTS AND DISCUSSION

4.1. Comparison with other existing methods

Table 5 shows the performance comparisons of proposed detection, segmentation and classification method with other existing methods as Sertan Kaymak et al. [22], Than Than Htay et al. [17], Ardalan Ghazemzadeh et al. [23], Simon Hadush et al. [24], and Tsehay Admassu Assegie et al. [25] (Figure 19).

The proposed method in this project obtained 96.67% of accuracy, 97.0% of precision, 96.0% of sensitivity, and 97.98 % of specificity (Table 4). The proposed method is based on GMM segmentation, GLCM for features extraction and PNN for classification. This proposed method is MATLAB GUI based model (Figure 18).

Two types of artificial neural networks been used in proposed method by Sertan Kaymak et al. [22] in their research. The classification in that method is obtained using "Radial Basis Neural Network" (RBFN) and "Back Propagation Neural Network" (BPPN). The accuracies of the BPNN and RBFN networks are 59.0% and 70.4% respectively. They used discrete Haar Wavelets method to improve the image quality.

The proposed method by Than Than Htay et al. [17] achieved 92.0% of accuracy in classify the type of tumor. The proposed method is using GLCM for feature extraction and K-Nearest Neighbor (KNN) classifier for detection of Lung cancer.

The method proposed by Ardalan Ghasemzadeh et al. [23] in their research had obtained 93.9% of accuracy, 95.1% of sensitivity, and 92.0% of specificity. Their method is based on Gabor-wavelet transform and SVM for classification of masses in mammograms.

The method projected by Simon Hadush et al. [24] obtained 91.86% of accuracy and 94.67% of sensitivity. The proposed method is using CNN for the bread cancer detection.

The method proposed by Tsehay Admassu et al. [25] in their work had obtained accuracy of 92.53%. Their method is based on adaptive boosting by using adaboost algorithm and used kaggle Lung cancer data repository.

Table 5: Result Performance of Previous Methods with Comparison to Existing Methods.

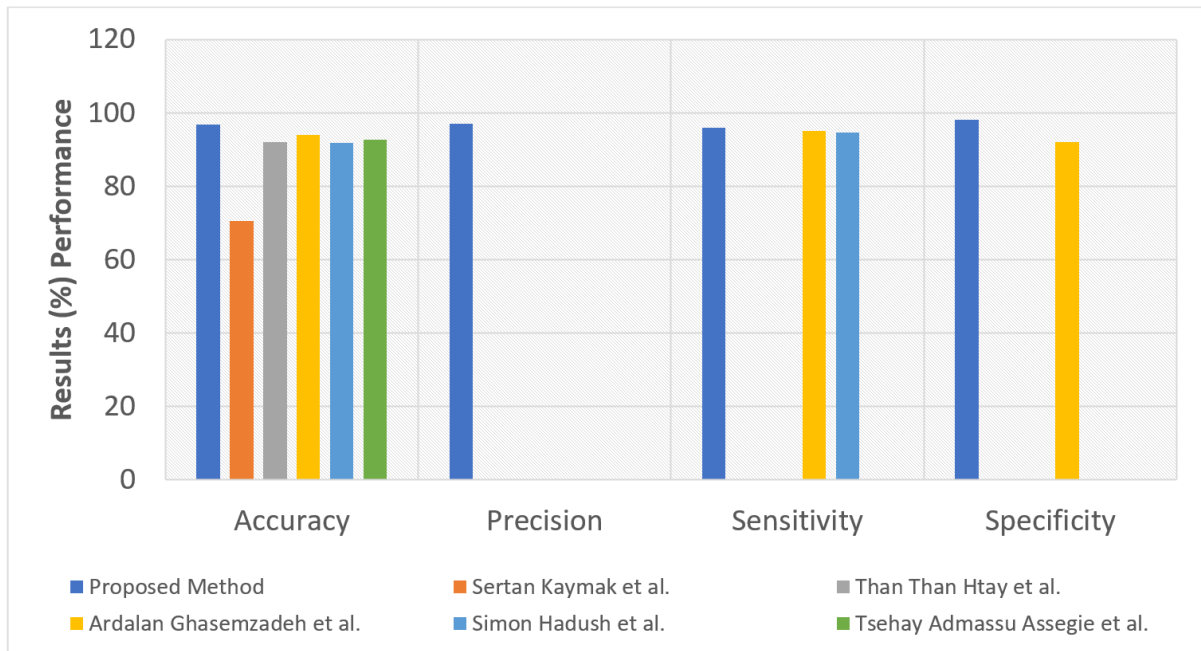| Results (%) Performance						
Evaluation Parameters	**Proposed Method**	**Sertan Kaymak et al. [22]**	**Than Than Htay et al. [17]**	**Ardalan Ghasemzadeh et al. [23]**	**Simon Hadush et al. [24]**	**Tsehay Admassu Assegie et al. [25]**
Accuracy	96.67	70.4	92.0	93.9	91.86	92.53
Precision	97.0	-	-	-	-	-
Sensitivity	96.0	-	-	95.1	94.67	-
Specificity	97.98	-	-	92.0	-	-

Figure 19: Result Performance of Previous Methods with Comparison to Existing Methods .

5. CONCLUSION

Lung cancer is amongst most common cancers diagnose by women/men every year globally and even the mortality rate due this cancer quite high compare to other cancers in women/men. The number of cancers can be reduced by detect and diagnose the disease at early stage. The available conventional methods like mammography screening have been one the most common detection method among Lung specialists. Due to some limitations, the accuracy of obtained data is low and led to some errors to occur. This situation had created a need to develop automated system to increase the accuracy rate.

In this project, an efficient method for segmentation and detection of Lung cancer utilizing artificial neural network classification method in the mammogram pictures been proposed. *'LungcancerGUI'* – A MATLAB based Graphical User Interface (GUI) is created in this project and it works as computer aided detection system in the proposed methodology. The images of mammogram used in this project are from mini-MIAS database. The mammogram images are uploaded to the *'LungcancerGUI'* and the image is pre-processed using adaptive median filtering method to remove noises and improve the quality. Then, the propose method is continued with segmentation process using Gaussian mixture model (GMM) method on the pre-processed image to retrieve the ROI from the mammogram image. The gray level co-occurrence matrix is also used to retrieve the characteristics of the mammography picture.

These extracted features are trained and classified using PNN classification approach. The PNN classifier identified the type of tumor and classified the mammogram into benign tumor, malignant tumor or normal type. The proposed method in this project obtained 96.67% of accuracy, 97.0% of precision, 96.0% of sensitivity, and 97.98% of specificity. There's always room for improvement and advancement in the Lung cancer detection by using neural networks and machine learning methods.

- Future work will concentrate on improving each stage from pre-processing till classification of Lung cancer by using more advance methods.

- There's still need some improvement in the graphical user interface in term of design and also its efficiency as a CAD system in the tumor detection process. Add up more features in the GUI.
- Concentrate on implement new neural network design/algorithm which can obtain higher accuracy than current ones.
- Use different modalities from other open access databases like DDSM, Image Retrieval in Medical Applications (IRMA) and Bayesian Decision Support in Medical Screening (B-SCREEN).
- This research work will be developed for the detection of severity level and also more different types of Lung cancer.

REFERENCES:

[1] S. Loibl, P. Poortmans, M. Morrow, C. Denkert, and G. Curigliano, "Lung cancer," *The Lancet*. 2021, doi: 10.1016/S0140-6736(20)32381-3.

[2] A. G. Waks and E. P. Winer, "Lung Cancer Treatment: A Review," *JAMA - J. Am. Med. Assoc.*, vol. 321, no. 3, pp. 288–300, 2019, doi: 10.1001/jama.2018.19323.

[3] A. L. Bane, S. Tjan, R. K. Parkes, I. Andrulis, and F. P. O'Malley, "Invasive lobular carcinoma: To grade or not to grade," *Mod. Pathol.*, 2005, doi: 10.1038/modpathol.3800273.

[4] G. Oda *et al.*, "Predictors for upstaging of ductal carcinoma in situ (DCIS) to invasive carcinoma in non-mass-type DCIS," *Mol. Clin. Oncol.*, 2020, doi: 10.3892/mco.2020.2036.

[5] S. Zhou *et al.*, "Effects of Lung Cancer Genes 1 and 2 on Cardiovascular Diseases," *Current Problems in Cardiology*. 2021, doi: 10.1016/j.cpcardiol.2019.04.001.

[6] F. Sardanelli *et al.*, "Mammography: an update of the EUSOBI recommendations on information for women/men," *Insights Imaging*, 2017, doi: 10.1007/s13244-016-0531-4.

[7] E. A. Sickles, "Screening for Lung cancer with mammography," *Clinical Imaging*. 1991, doi: 10.1016/s0140-6736(01)07198-7.

[8] A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, and E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," *Proc. Natl. Acad. Sci. U. S. A.*, 2020, doi: 10.1073/pnas.1919012117.

[9] A. Chien, J. S. Weaver, E. Kinne, and I. Omar, "Magnetic resonance imaging of the knee," *Polish J. Radiol.*, 2020, doi: 10.5114/pjr.2020.99415.

[10] S. M. McKinney *et al.*, "International evaluation of an AI system for Lung cancer screening," *Nature*, 2020, doi: 10.1038/s41586-019-1799-6.

[11] R. Ramani, N. S. Vanitha, and S. Valarmathy, "The Pre-Processing Techniques for Lung Cancer Detection in Mammography Images," *Int. J. Image, Graph. Signal Process.*, 2013, doi: 10.5815/ijigsp.2013.05.06.

[12] A. K. Singh and B. Gupta, "A Novel Approach for Lung Cancer Detection and Segmentation in a Mammogram," 2015, doi: 10.1016/j.procs.2015.06.079.

[13] H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, "Automated Lung cancer detection and classification using ultrasound images: A survey," *Pattern Recognit.*, 2010, doi: 10.1016/j.patcog.2009.05.012.

[14] N. Kaur and A. S. Sharma, "Detection of Lung cancer on mammograms using neural network approach," 2019, doi: 10.1109/GUCON.2018.8675017.

[15] E. Y. K. Ng and E. C. Kee, "Advanced integrated technique in Lung cancer thermography," *J. Med. Eng. Technol.*, 2008, doi: 10.1080/03091900600562040.

[16] A. Elmoufidi, "Pre-processing algorithms on digital X-ray mammograms," 2019, doi: 10.1109/ISC246665.2019.9071750.

[17] T. T. Htay and S. S. Maung, "Early Stage Lung Cancer Detection System using GLCM feature extraction and K-Nearest Neighbor (k-NN) on Mammography image," 2018, doi: 10.1109/ISCIT.2018.8587920.

[18]   S. Shamy and J. Dheeba, "A research on detection and classification of Lung cancer using k-means gmm & CNN algorithms," *Int. J. Eng. Adv. Technol.*, 2019, doi: 10.35940/ijeat.F1102.0886S19.

[19]   A. Y. Ebrahim, "Detection of Lung cancer in mammograms through a new features and decision tree based, classification framework," *J. Theor. Appl. Inf. Technol.*, 2017.

[20]   I. S. Keita, I. P. Nair, H. Prasad, S. Pandian, and S. Deivasigamani, "Classification of Benign and Malignant MRIs using SVM Classifier for Brain Tumor Detection," *Int. J. Eng. Trends Technol.*, vol. 70, no. 2, pp. 221–228, 2022, doi: 10.14445/22315381/IJETT-V70I2P226.

[21]   D. S., V. Ganesan, D. Pazhani, M. Ramasamy, and U. -, "Gabor Transform Based Glioma Brain Tumor Detection Using Neural Networks," *Int. J. Eng. Technol.*, 2018, doi: 10.21817/ijet/2018/v10i6/181006008.

[22]   S. Kaymak, A. Helwan, and D. Uzun, "Lung cancer image classification using artificial neural networks," 2017, doi: 10.1016/j.procs.2017.11.219.

[23]   A. Ghasemzadeh, S. Sarbazi Azad, and E. Esmaeili, "Lung cancer detection based on Gabor-wavelet transform and machine learning methods," *Int. J. Mach. Learn. Cybern.*, 2019, doi: 10.1007/s13042-018-0837-2.

[24]   S. A. Alanazi *et al.*, "Boosting Lung Cancer Detection Using Convolutional Neural Network," *J. Healthc. Eng.*, 2021, doi: 10.1155/2021/5528622.

[25]   T. A. Assegie, R. L. Tulasi, and N. K. Kumar, "Lung cancer prediction model with decision tree and adaptive boosting," *IAES Int. J. Artif. Intell.*, 2021, doi: 10.11591/ijai.v10.i1.pp184-190.